**Overview of the Pilot Performance Dataset and Analysis**

**Data Description:**

The dataset consists of performance metrics collected from commercial airline pilots during a simulated Airbus A320 test environment. The dataset includes 11 columns and contains data on 11,050 recorded scenarios. The fields in the dataset are as follows:

- Pilot ID: Unique identifier for each pilot.
- Scenario: The type of simulation scenario the pilot was subjected to (e.g., Turbulence).
- Cockpit Item: Specific cockpit system or component being evaluated (e.g., Primary Flight Display, Autopilot System).
- Reaction Time (s): The time taken by the pilot to react to a stimulus, measured in seconds.
- Correct Action?: A binary indicator of whether the pilot performed the correct action (Yes or No).
- Blood Pressure (systolic): The systolic blood pressure of the pilot during the scenario, measured in mmHg.
- Blood Pressure (diastolic): The diastolic blood pressure of the pilot during the scenario, measured in mmHg.
- Heart Rate: The pilot's heart rate in beats per minute (bpm).
- Oxygen Saturation: The pilot's blood oxygen saturation level, expressed as a percentage.
- Tremor Intensity: A quantitative measure of hand tremor intensity, reflecting physical stress levels.

- Perspiration: A measure of perspiration during the scenario, as a marker of stress.
- Purpose of the Dataset

This dataset was designed to evaluate and quantify the response rates and decision-making abilities of commercial airline pilots under varying stress levels. The goal was to simulate real-world challenges and assess:

- Reaction times under stress.
- Physiological responses, such as heart rate, blood pressure, and oxygen saturation.
- Decision-making accuracy, reflected by the "Correct Action?" field.
- How external factors, such as specific cockpit components, influence performance.

**Scenarios:**

The simulations placed pilots in various controlled but rigorous scenarios to test their cognitive and physical responses. Example scenarios include:

- Light Turbulence: Minimal stress situation to establish baseline measurements.
- Severe Turbulence: High-stress environment to observe performance under extreme conditions.
- System Failures: Testing decision-making when primary systems, such as navigation or autopilot, become inoperative.
- Emergency Landings: Assessing reaction times and decision accuracy in critical situations.

- Machine Learning and Predictive Models
- Using this dataset, various predictive models can be developed to:
- Classify Correct Actions: Determine the likelihood of a correct decision based on physiological and scenario-based features.
- Predict Reaction Time: Use regression models to predict a pilot's reaction time under different conditions.
- Identify Key Stress Indicators: Analyze which physiological markers (e.g., tremor intensity, heart rate) are most predictive of performance outcomes.
- Assess Training Needs: Identify pilots who may require additional training based on model predictions.
- Methodology

The analysis included the following steps:

**Data Preprocessing:**

- Categorical fields (e.g., Scenario, Cockpit Item) were encoded for machine learning models.
- Numerical features were standardized to ensure uniform scaling.

**Model Selection:**

- A Random Forest Classifier was employed to predict whether a pilot performed the correct action based on input features.
- Model performance was evaluated using metrics such as precision, recall, and F1 score.

**Feature Importance:**

- The most impactful features influencing pilot performance were identified (e.g., Reaction Time, Tremor Intensity).

**Data Insights:**

- Visualizations and statistics provided actionable insights into physiological stress and performance trends.

**Conclusion:**

This dataset serves as a comprehensive tool for analyzing the responses of commercial airline pilots in a simulated Airbus A320 environment. By subjecting pilots to rigorous scenarios, the study aimed to enhance training protocols, improve safety, and refine pilot performance metrics. The integration of machine learning models further enables targeted analysis, offering data-driven recommendations to optimize pilot training and safety standards.

This document provides an overview of the dataset and its analytical potential. For further details, refer to the MATLAB code included in this repository, which demonstrates the preprocessing steps, model training, and feature analysis.